Introduction

Welcome to PyRec! This Python-based application was designed to assist in the generation of DNA oligonucleotides for recombination. PyRec 3.0 generates oligonucleotides for generating stop codons in combination with a unique restriction site that can be used to make null mutations in genes. In addition the program generates oligonucleotides for screening for the incorporation of the mutations generated by recombineering.

Before You Use PyRec

Installations required to run PyRec – Windows

Python

Python can be downloaded for Windows at http://www.python.org/download/releases/2.7/. It is critical that version of Python is 2.7 as PyRec is not compatible with Python 3.2. You can download Python using the following link. Windows x86 MSI Installer (2.7) (sig).

Follow the download instructions provided:

1. Select "Install for all users", click next
2. Keep the default directory, click next
3. Customize: do not change(e.g., install all packages), click next
4. Install, press finish

Biopython

PyRec also requires the Biopython module. This is available for free download. Use the following link to download BioPython that is compatible with Python 2.7. http://biopython.org/DIST/biopython-1.55.win32-py2.7.exe.

Follow the download instructions:

1. Run the installer, it should detect Python automatically, click next
2. Keep the default directory, click next
3. Click next again to install

Adding python to the command line

Once you have installed python, you must configure your command line to recognize "python" as a command. Here is a step-by-step guide to achieving this[i]:

Directions for Windows XP

1. Right-click "My Computer" and select "Properties"
2. Under the "Advanced" tab, click "Environment Variables" (button towards the bottom)
3. In the "System Variables" box, find the variable named "Path"

4. If “Path” does not exist, click “New”
    a. In the “Variable Name” box, type “PATH” (omit the quotes)
    b. In the “Variable Value” box, type “%PATH%;C:\Python27” (omit the quotes)
    c. Click “OK” on each window you opened to save your changes
5. If “Path” already exists,
    a. Highlight “Path” and click “Edit”
    b. There is an entry field labeled “Variable Value”. Place your cursor at the very end of the field. WARNING: Make sure your cursor is at the very end of the field and you are not modifying any existing text. Type “;C:\Python27” (note the semicolon, and omit the quotes)
    c. Click “OK” on each window you opened to save your changes

Directions for Windows 7

1. Click on the circular start menu icon and search for “Environment Variables”
2. Select “Edit the system environment variables”
3. Select “Environment Variables” (button towards the bottom)
4. In the “System Variables” box, find the variable named “Path”
5. If “Path” does not exist, click “New”
    a. In the “Variable Name” box, type “PATH” (omit the quotes)
    b. In the “Variable Value” box, type “%PATH%;C:\Python27” (omit the quotes)
    c. Click “OK” on each window you opened to save your changes
6. If “Path” already exists,
    a. Highlight “Path” and click “Edit”
    b. There is an entry field labeled “Variable Value”. Place your cursor at the very end of the field. WARNING: Make sure your cursor is at the very end of the field and you are not modifying any existing text. Type “;C:\Python27” (note the semicolon, and omit the quotes)
    c. Click “OK” on each window you opened to save your changes

How to Use PyRec

To install PyRec 3.0 on your machine double click the PyRec.zip file to unzip and a PyRec file will be placed in the same directory you unzipped the file. The following files will be extracted:

Files - contains all the Python scripts that make up PyRec – don’t touch!

Genes – an empty directory that is where the program looks for the genes (fasta format) you wish to generate oligos for recombineering.

Oligos – an empty directory where the output oligos will be stored (csv format).

Change log.txt – describes changes made from the inception of PyRec.

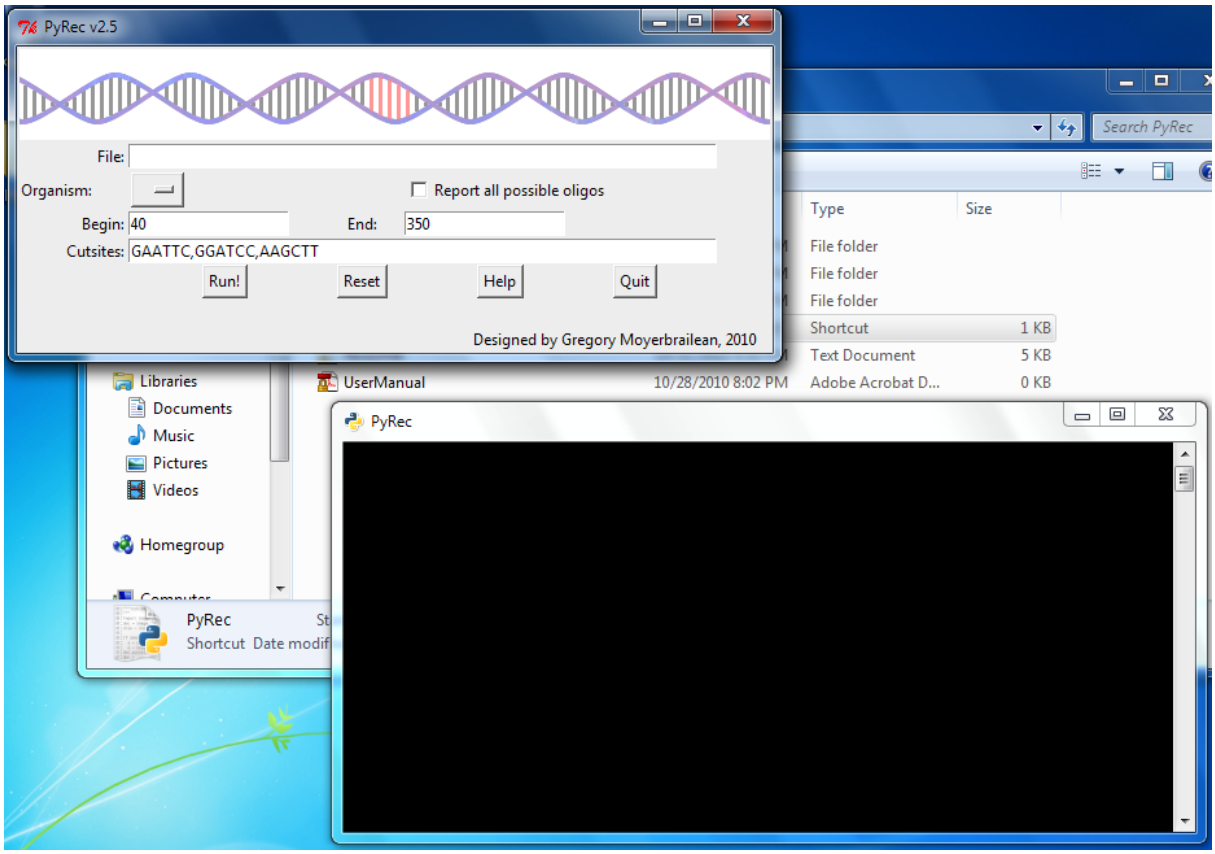
PyRec.py – The executable version of PyRec.

PyRec User Manual.

Getting started

The gene to be analyzed by PyRec must be placed in a fasta format in a simple text editor such as NotePad. Once the gene sequence is listed in the fasta format it must be saved to the Genes directory in the PyRec folder as a .fasta file (eg. rbgA.fasta). One fasta file can also contain multiple genes in fasta format. Each will be processed separately and separate output files will be generated for each individual gene.

Double click the PyRec.py file to open up the PyRec interface. Two windows will open – one is labeled PyRec and is the window in which all commands will take place. The other is a window that shows feedback on how the program is progressing during the current run.



To run the script on a particular file of one or more genes enter the filename in the File window. For example - genes/rbgA.fasta. Specify the organism in the drop down menu (only *L. lactis* in this version).

Click Run. Depending on the number of genes in your file, this may take several seconds to several minutes. If your parameters are not properly set, a popup will inform you of your error. If a critical error occurs within the program, the error will report on the console screen. If the program runs without error, the console will report that the oligo generation is complete.

PyRec will scan each gene from nucleotides 40 to 350 to identify genes that can be modified to generate an in-frame stop codon and a unique restriction site by making five adjacent mismatches in the coding sequence. Three restriction sites are analyzed – EcoRI, HindIII, and BamHI. One recombineering oligonucleotide (90 mer) will be reported for each restriction site (if one can be made). In addition, three additional oligonucleotides accompanying the recombineering oligonucleotide that can be used for screening for mutations are output. A forward primer located ~500bp upstream of the mutations is generated, a reverse primer located ~500 bp downstream of the mutations is generated, and a primer for performing MAMA-PCR (mismatch amplification mutation assay PCR) is generated, to be used in combination with the forward primer. There are two methodologies for screening. The forward and reverse primers will amplify a ~1kB fragment that can be analyzed by restriction digest. Only those cells that have incorporated the mutation will yield a partial digest. As an alternative method the forward primer can be used with the MAMA primer to generate a 500 bp amplicon only if the mutation has been incorporated into the genome.

Examples of rbgA oligonucleotides that would be produced.

Recombineering oligonucleotide – rbgA_E. rbgA is the gene and E refers to EcoRI as the restriction site. H for HindIII, B for BamHI. If you decide to use a restriction site other than these, the oligo will be named with the six nucleotide cut site sequence.

Forward primer – rbgA_fwd_E

Reverse primer – rbgA_rev_E

MAMA primer – rbgA_MAMA_E

Note: All oligos are oriented 5’ – 3’.

Output files

A .csv file that can be opened in Excel is generated for each gene and placed in the “L.lactismg1363” folder that resides in the “Oligos” folder located in the “PyRec” folder. Open the file and the following worksheet will appear:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Oligo and primers designed using the recombineering module for python.															
2	Written and designed by Gregory Moyerbrailean.															
3																
4																
5	Gene Information:															
6	ID:	rbgA														
7	Description:	rbgA														
8	Length:	1365														
9																
10	BLAST Information:															
11	Match:	No BLAST data														
12	E-Value:	No BLAST data														
13	Location	80														
14	Direction	forward														
15	Position	Early														
16																
17																
18	Oligo Designs:															
19		Mutation made at nt: 68														
20	Oligo	Sequence		Forward	Sequence	Length	T-anneal	MAMA	Sequence	Length	T-anneal	Reverse	Sequence	Length	T-anneal	Product
21	rbgA_B	acagacattaattgagcaggctcar		rbgA_fwd_B	cttttcttttagagt	30	59.9333333	rbgA_mama_	ctaggcagagta	21	59.8238095	rbgA_rev_B	ctttccatctgattt	26	60.0692308	1079
22																
23																
24		Mutation made at nt: 96														
25	Oligo	Sequence		Forward	Sequence	Length	T-anneal	MAMA	Sequence	Length	T-anneal	Reverse	Sequence	Length	T-anneal	Product
26	rbgA_E	tcaagcaagatattggcagtatctttg		rbgA_fwd_E	gtagtactcttttc	30	58.5666667	rbgA_mama_	acaagcttatga	30	57.2	rbgA_rev_E	aagtaagtctaa	30	59.9333333	1074
27																

As an example, let us consider the recombineering oligonucleotide rbgA_B ("B" designating the creation of a BamHI restriction site). Mutation made at nt 68 refers to the first bp that is mutated by the recombineering oligonucleotide. There are always five bp mismatches in a row. The recombineering oligonucleotide – name under Oligo (column A) and sequence under Sequence (column B). The forward oligonucleotide name is listed under Forward (column D) along with sequence (column E), length (column F), and the predicted Tm of the oligo (Column G). The same outputs are listed for the MAMA-PCR primer and the Reverse Primer. In column P the amplicon size using the forward and reverse primers is noted.

A note about the output: PyRec has a feature that will check the wildtype sequence surrounding the mutation for similarities. Such similar sequences were implicated in false positive results in oligos with three adjacent mutations, but it is yet unknown how such similarities affect oligos with five adjacent mutations. PyRec therefore reserves a spot in the output (column C) where it will make note if a similar sequence has been detected.

New in version 3.1: PyRec 3.1 includes a new feature that will warn you if the gene you ran is present multiple times within the genome. PyRec will always generate screening oligos based off of the first copy it finds, starting at the first nucleotide of the (+)-sense genomic sequence and continuing to the last nucleotide of the (-)-sense genomic sequence. At this time, there is no way to specify for which copy of the gene to create screening oligos.

Options

The default for the program attempts to identify stop codons in the 5' end of the gene to optimize the chance of generating a null mutation. The following are the defaults for PyRec and how they can be modified as desired by the user.

Begin and **End**: Specify the section of the gene to search for mutation sites. Numbers represent nucleotide positions within the gene. The default is nt 40-350 and can me modified by the user as desired.

Cutsites: Specify the cutsites of the restriction enzymes you wish to use. The defaults are EcoRI, BamHI, and HindIII. You may edit this list as desired by adding the actual 6bp site in the cutsite window separated by commas. Note that at this time PyRec only supports 6bp cutsites.

Report all possible oligos: By default, PyRec will restrict your output to one oligo per cutsite specified. Check this option to report all oligos the program generates.

[i] Adapted from "http://www.voidspace.org.uk/python/articles/command_line.shtml#environment-variables".